UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/776,892	02/11/2004	Kirill Stoimenov	9432-000257	3341

27572 7590 06/07/2007
HARNESS, DICKEY & PIERCE, P.L.C.
P.O. BOX 828
BLOOMFIELD HILLS, MI 48303

EXAMINER	
SIEDLER, DOROTHY S	

ART UNIT	PAPER NUMBER
2626	

MAIL DATE	DELIVERY MODE
06/07/2007	PAPER

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Office Action Summary	Application No.	Applicant(s)	
	10/776,892	STOIMENOV ET AL.	
	Examiner	Art Unit	
	Dorothy Sarah Siedler	2626	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE 3 MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on *11 February 2004*.

2a)☐ This action is **FINAL**. 2b)☒ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) *1-28* is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) *1-28* is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☒ The drawing(s) filed on *11 February 2004* is/are: a)☒ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☐ All b)☐ Some * c)☐ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☒ Notice of References Cited (PTO-892)
2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)
3) ☒ Information Disclosure Statement(s) (PTO/SB/08) Paper No(s)/Mail Date *2-11-04*.
4) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____ .
5) ☐ Notice of Informal Patent Application
6) ☐ Other: _____.

DETAILED ACTION

This is the initial response to the application filled February 11, 2004. Claims 1-28 are pending and are considered below.

Claim Rejections - 35 USC § 112

The following is a quotation of the second paragraph of 35 U.S.C. 112:

> The specification shall conclude with one or more claims particularly pointing out and distinctly claiming the subject matter which the applicant regards as his invention.

Claims 10,14- 17,25-28 are rejected under 35 U.S.C. 112, second paragraph, as being indefinite for failing to particularly point out and distinctly claim the subject matter which applicant regards as the invention.

Claim 10 recites the limitation "said text segment". There is insufficient antecedent basis for this limitation in the claim.

Claim 10 recites, "wherein said input-text contains at least one non-editable said text segment and at least one editable said segment", however this is ambiguous. It is unclear whether "said text segment" and "said segment" refer to the same segment, or distinct segments. In addition, the "said segment", as defined in claim 1, is a segment from the processed representation, which is created from the input-text by a text-to-speech engine. Therefore the input-text cannot contain a segment, since a segment is not created until the input-text is processed by the text-to-speech engine. Therefore the examiner interprets claim 10 as the visual editing interface allowing editing of the segment, wherein the segment is the text-input converted by the text-to-speech engine. This interpretation used throughout the remainder of this office action.

Claims 14 and 15 recite, "wherein said processed representation is a textual representation" and "wherein said textual representation is used to generate said processed representation", however this is ambiguous. Claims 14 and 15 can be interpreted as, "said processed representation is a textual representation used to generate the processed representation" or the processed representation is used to generate itself. Since this cannot logically be true, the examiner interprets claim 14 as claiming wherein the processed representation is a modified textual representation of the processed input text. Claim 15 is interpreted as claiming wherein the input-text is used to generate the processed representation. These interpretations are used to interpret claims 16 and 17 as well. Claim 16 is interpreted as claiming wherein said modified textual representation (or the processed representation) is stored and accessed from a data store. Claim 17 is interpreted as claiming the modified textual representation (or the processed representation) is used to generate synthesized speech using a TTS system. These interpretations used throughout the remainder of this office action.

Claims 25 and 26 are similar to claims 14 and 15, and are therefore rejected for the same reasons and interpreted the same. Claims 27 and 28 are similar to claims 16 and 17, and are therefore rejected for the same reasons an interpreted the same.

Claim Rejections - 35 USC § 103

The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

Claims 1-28 are rejected under 35 U.S.C. 103(a) as being unpatentable over ***Taylor*** ("SSML: A Speech Synthesis Markup Language" Speech Communication, 1996) in view of ***Kobal*** (7,099,828).

As per claim 1, ***Taylor*** discloses a system for tuning the text-to-speech conversion process, the system comprising: a text-to-speech engine, said text-to-speech engine receiving at least one text-input and converting said text-input into a processed representation, said processed representation including at least one speech feature associated with at least one segment of said representation (page 3, Section 1.1 Annotated Text in Speech Synthesis, *a markup language is used to annotate, or tag, input text (processed representation), the tags indicating a pronunciation of the input text word or phrase*). ***Taylor*** does not explicitly disclose a visual editing interface, said visual editing interface displaying said processed representation using at least one graphical indicator on an output device, wherein said segment is displayed on said output device using said graphical indicator corresponding to said speech feature. However ***Taylor*** does disclose that most SGML documents, such as HTML, are

physically typed at keyboards (page 17, Section 5.2, first paragraph). This implies the presence of a word processor, enabling a user to enter the text, including tags, and edit the SGML document. ***Taylor*** also discloses the use of a level tag, which is used to indicate the amount of automatic prosodic analysis initially performed by a machine (pages 12-13, Section 3.4). The level tag enables a user to indicate when the system should automatically produce prosodic tags, and when they should be provided by the user, for example through editing. In addition, ***Kobal*** discloses a graphical user interface, which presents a visual identifier (graphical indicator) corresponding to a phoneme (speech feature) used to indicate the pronunciation of text in a text to speech system (Abstract and Figure 2). The user interface enable enables the user to adjust pronunciation features, including prosody and stress, using a graphical tool, such as buttons. ***Taylor*** and ***Kobal*** both disclose systems for the adjustment of prosodic features used during speech synthesis, and are therefore analogous art.

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to have a visual editing interface displaying said processed representation using at least one graphical indicator on an output device, wherein said segment is displayed on said output device using said graphical indicator corresponding to said speech feature in ***Taylor***, since a graphical indicator provides an simple and reliable method for users to adjust speech features of text, including pronunciation features, as indicated in ***Kobol*** (column 1 lines 50-60).

As per claim 18, ***Taylor*** does not explicitly disclose a system for providing a text-to-speech interface, the system comprising: a visual interface connected to a text-to-speech engine; and at least one communication channel connecting said visual interface to said text-to-speech engine, said text-to-speech engine communicating with said visual interface over said communication channel by sending and receiving at least one data segment in a format. However ***Taylor*** does disclose that most SGML documents, such as HTML, are physically typed at keyboards (page 17, Section 5.2, first paragraph). This implies the presence of a word processor, enabling a user to enter the text, including tags, and edit the SGML document. ***Taylor*** also discloses the use of a level tag, which is used to indicate the amount of automatic prosodic analysis initially performed by a machine (pages 12-13, Section 3.4). The level tag enables a user to indicate when the system should automatically produce prosodic tags, and when they should be provided by the user, for example through editing. In addition, ***Kobal*** discloses a graphical user interface, which communicates with a pronunciation processor to send and receive data (column 4 lines 11-33 and Figure 1). ***Taylor*** and ***Kobal*** both disclose systems for the adjustment of prosodic features used during speech synthesis, and are therefore analogous art.

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to have a visual interface connected to a text-to-speech engine, and at least one communication channel connecting said visual interface to said text-to-speech engine, said text-to-speech engine communicating with said visual interface over said communication channel by sending and receiving at least one data segment in a format

in ***Taylor***, since it would enable the system to respond to requests received from the user, as indicated in ***Kobal*** (colum 4 lines 24-25).

As per claim 22, ***Taylor*** disclose a method for visual tuning text-to-speech conversion process, the method comprising: converting an input-text to a processed representation using a text-to-speech engine, said processed representation including at least one speech feature of said input-text (page 3, Section 1.1 Annotated Text in Speech Synthesis, *a markup language is used to annotate, or tag, input text (processed representation), the tags indicating a pronunciation of the input text word or phrase*); ***Taylor*** does not explicitly disclose displaying said processed representation on a visual editing interface connected to said text-to-speech engine, said speech feature of said processed representation being displayed in a corresponding graphical form, and providing an editing function in said visual editing interface to a user for modifying said speech feature in said graphical form. However ***Taylor*** does disclose that most SGML documents, such as HTML, are physically typed at keyboards (page 17, Section 5.2, first paragraph). This implies the presence of a word processor, enabling a user to enter the text, including tags, and edit the SGML document. ***Taylor*** also discloses the use of a level tag, which is used to indicate the amount of automatic prosodic analysis initially performed by a machine (pages 12-13, Section 3.4). The level tag enables a user to indicate when the system should automatically produce prosodic tags, and when they should be provided by the user, for example through editing. In addition, ***Kobal*** discloses a graphical user interface, which presents a visual identifier (graphical

indicator) corresponding to a phoneme (speech feature) used to indicate the pronunciation of text in a text to speech system (Abstract and Figure 2). The user interface enables the user to adjust pronunciation features, including prosody and stress, using a graphical tool, such as buttons. ***Taylor*** and ***Kobal*** both disclose systems for the adjustment of prosodic features used during speech synthesis, and are therefore analogous art.

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to display said processed representation on a visual editing interface connected to said text-to-speech engine, said speech feature of said processed representation being displayed in a corresponding graphical form, and providing an editing function in said visual editing interface to a user for modifying said speech feature in said graphical form in ***Taylor***, since a graphical indicator provides an simple and reliable method for users to adjust speech features of text, including pronunciation features, as indicated in ***Kobol*** (column 1 lines 50-60).

As per claim 2, ***Taylor*** in view of ***Kobol*** disclose the system of claim 1, and ***Kobol*** further discloses wherein said visual editing interface provides at least one editing function to a user, the editing function enabling the modification of said speech feature associated with said segment through a change in the corresponding said graphical indicator (column 6 lines 12-30, *visual identifiers (graphical indicators), such as buttons,*

are used to change corresponding pronunciation features, including prosodics, (speech feature) of the text).

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to have a visual editing interface that provides at least one editing function to the user, the editing function enabling the modification of said speech feature associated with said segment through a change in the corresponding said graphical indicator in ***Taylor***, since a graphical indicator provides an simple and reliable method for users to adjust speech features of text, including pronunciation features, as indicated in ***Kobol*** (column 1 lines 50-60).

As per claim 3, ***Taylor*** in view of ***Kobol*** disclose the system of claim 2, and ***Kobol*** further discloses a visual editing interface that associates said speech feature corresponding to said segment with said graphical indicator, wherein the user's modification of said graphical indicator results in a corresponding change in said speech feature of said segment (column 6 lines 12-30, *visual identifiers (graphical indicators), such as buttons, are used to change corresponding pronunciation features, including prosodics, (speech feature) of the text).*

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to a visual editing interface that associates said speech feature corresponding to said segment with said graphical indicator, wherein the user's modification of said graphical indicator results in a corresponding change in said speech

feature of said segment in ***Taylor***, since a graphical indicator provides an simple and reliable method for users to adjust speech features of text, including pronunciation features, as indicated in ***Kobol*** (column 1 lines 50-60).

As per claim 4, ***Taylor*** in view of ***Kobol*** disclose the system of claim 1, and ***Taylor*** further discloses wherein said speech feature is at least one of the following: normalized text, part-of-speech, parsing of text, chunking of text, boundary strength, pause duration, transcription, speech rate, syllable duration, segment duration, pitch, word prominence, emphasis, formant mixing mode, unit selection override, intensity contour, formant trajectories, and allophone rules (page 9-11, Section 3.2 Set of Example Tags, *the Speech Synthesis Markup Language, used to annotate the text, includes tags indicating intonational phrase boundary and emphasis*) .

As per claim 5, ***Taylor*** in view of ***Kobol*** disclose the system of claim 1, and ***Kobol*** further discloses wherein said graphical indicator comprises at least one of the following: graphical style, font faces, coloring, vertical spacing, horizontal spacing, italicization, boldness, underlining, blinking, crossing-out, text orientation, text rotation, punctuation symbols and graphical symbols (column 5 lines 12-30 and Figure 2, *the pronunciation being composed is represented by a text (graphical style) corresponding to the phoneme to be pronounced*).

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to have a graphical indicator include a graphical style in ***Taylor***, since a

graphical representation, or a text representation, is a common representation used by editing systems which can be easily read and understood by a user.

As per claims 6 and 19, ***Taylor*** in view of ***Kobol*** disclose the system of claims 1 and 18, and ***Taylor*** further discloses wherein said processed representation employs a parameterized aligned sound records format (page 19 and 20, *examples of the SSML tags and text used are provided, which are equivalent to the format style of parameterized aligned sound records format*).

As per claim 7, ***Taylor*** in view of ***Kobol*** disclose the system of claim 1, and ***Taylor*** further discloses wherein said segment comprises at least one of the following: word, letter, syllable, pause, word boundary and punctuation-mark (Page 9-11, *tags are used in association with a phrase or word*).

As per claims 8 and 9, ***Taylor*** in view of ***Kobol*** disclose the system of claim 1, however neither explicitly disclose wherein said visual editing interface operates as a plug-in for a graphical user interface, wherein said plug-in is an ActiveX control. However, ***Kobol*** does disclose that system can be used as a standalone tool, or it can be included in a larger application (column 3 lines 46-49). In addition, Active-x controls were developed in the 1990's by Microsoft to enable enhanced formatting of web pages. Using the standard HTML <object> tags, Active-x enables the users to specify data to control, and how to control it, enabling the web a page to behave more like a program than static

pages. Therefore the examiner argues that the use of Active-x controls are old and well known.

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to have a visual editing interface that operates as a plug-in control for a GUI, wherein the plug-in is an Active-X control in ***Taylor***, since it provides reliable, readily available software, removing the need to spend time and resources developing new software to the implementation of the system on a web page.

As per claim 10, ***Taylor*** in view of ***Kobol*** disclose the system of claim 1, however ***Taylor*** does not explicitly disclose wherein said visual editing interface allows editing of said input-text wherein said input-text contains at least one non-editable said text segment and at least one editable said segment. ***Taylor*** does disclose that most SGML documents, such as HTML, are physically typed at keyboards (page 17, Section 5.2, first paragraph). This implies the presence of a word processor, enabling a user to enter the text, including tags, and edit the SGML document. ***Taylor*** also discloses the use of a level tag, which is used to indicate the amount of automatic prosodic analysis initially performed by a machine (pages 12-13, Section 3.4). The level tag enables a user to indicate when the system should automatically produce prosodic tags, and when they should be provided by the user, for example through editing. In addition, ***Kobal*** discloses a graphical user interface, which presents a visual identifier (graphical indicator) corresponding to a phoneme (speech feature) used to indicate the pronunciation of text in a text to speech system (Abstract and Figure 2). The user

interface enable enables the user to adjust pronunciation features, including prosody and stress, using a graphical tool, such as buttons. ***Taylor*** and ***Kobal*** both disclose systems for the adjustment of prosodic features used during speech synthesis, and are therefore analogous art.

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to have a visual editing interface displaying said processed representation using at least one graphical indicator on an output device, wherein said segment is displayed on said output device using said graphical indicator corresponding to said speech feature in ***Taylor***, since a graphical indicator provides an simple and reliable method for users to adjust speech features of text, including pronunciation features, as indicated in ***Kobol*** (column 1 lines 50-60).

As per claim 11, ***Taylor*** in view of ***Kobol*** disclose the system of claim 1, and ***Kobol*** further discloses wherein said visual editing interface is language independent (column 5 lines 60-67).

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to have the visual interface be language independent in ***Taylor***, since it would enable the system to be used for applications in more than one language, as indicated in ***Kobol*** (column 3 lines 30-32).

As per claims 12 and 23, ***Taylor*** in view of ***Kobol*** disclose the system of claims 1 and 22, and ***Kobol*** further discloses wherein said visual editing interface provides the user with speech audio output of said processed representation (column 5 lines 36-39 and Figure 2, item 230).

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to have the visual display provide the user with speech audio output in ***Taylor***, since it would enable playback of the speech which is also displayed in the window, as indicated in ***Kobol*** (column 5 lines 36-39).

As per claims 14,15,25 and 26, ***Taylor*** in view of ***Kobol*** disclose the system of claims 1 and 22, and ***Taylor*** further discloses wherein the said processed representation is a textual representation, wherein the said textual representation is used to generate said processed representation (page 9, section 3.2 Set of Example Tags and page 19, Example SSML Documents, *input text is used to create SSML text documents*).

As per claims 13,16, 24 and 27 ***Taylor*** in view of ***Kobol*** disclose the system of claims 1 and 15, and ***Kobol*** further discloses wherein visual editing interface is connected to a data-store for storing and retrieving said representation (column 5 lines 42-43, *the pronunciation (representation) is saved, and can be opened for further editing*).

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to have a visual editing machine connected to a data-store for storing

and retrieving said presentation in ***Taylor***, since it would enable the user to create a presentation, then save the information for editing at a later time.

As per claims 17 and 28, ***Taylor*** in view of ***Kobol*** disclose the system of claims 14 and 25, and ***Taylor*** further discloses wherein said textual representation is used to generate synthesized speech using a TTS system distinct from said text-to-speech engine (page 14, sections 4.2 SSML Interpreter and 4.3. Synthesizer Operation, and Figure 2, *an SSML document is created, then passed to a synthesizer which outputs synthesized audio*)

As per claim 20, ***Taylor*** in view of ***Kobol*** disclose the system of claim 18, however ***Taylor*** does not disclose wherein said text-to-speech engine sends said data segment in the parameterized aligned sound records format to said visual interface, said visual interface rendering said data segment in a visual form, said visual interface allowing editing of said data segment to produce an edited data segment, said visual interface sending said edited data segment to said text-to-speech engine. ***Kobal*** discloses a graphical user interface, which presents a visual identifier (graphical indicator) corresponding to a phoneme (speech feature) used to indicate the pronunciation of text in a text to speech system (Abstract and Figure 2). The user interface enable enables the user to adjust pronunciation features, including prosody and stress, using a graphical tool, such as buttons. The user interface also communicates with a

pronunciation processor to send and receive data, including pronunciation information (column 4 lines 11-33 and Figure 1).

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to have a text-to-speech engine that sends said data segment in the parameterized aligned sound records format to said visual interface, said visual interface rendering said data segment in a visual form, said visual interface allowing editing of said data segment to produce an edited data segment, said visual interface sending said edited data segment to said text-to-speech engine in ***Taylor***, since a graphical indicator provides an simple and reliable method for users to adjust speech features of text, including pronunciation features, as indicated in ***Kobol*** (column 1 lines 50-60).

As per claim 21, ***Taylor*** in view of ***Kobol*** disclose the system of claim 18, however neither explicitly wherein said visual interface sends data to said text-to-speech engine over a first said communication channel and said text-to-speech engine sends data to said visual interface over a second said communication channel. However, ***Kobol*** does disclose a graphical user interface, which communicates with a pronunciation processor to send and receive data (column 4 lines 11-33 and Figure 1).

Therefore it would have been obvious to one of ordinary skill in the art at the time of the invention to have a visual interface send data to said text-to-speech engine over a

first said communication channel and have said text-to-speech engine send data to said visual interface over a second said communication channel in ***Taylor***,

Conclusion

The prior art made of record and not relied upon is considered pertinent to applicant's disclosure. Please see the PTO-892 form.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to Dorothy Sarah Siedler whose telephone number is 571-270-1067. The examiner can normally be reached on Mon-Thur 9:30am-5:30pm.

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Richemond Dorvil can be reached on 571-272-7602. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

DSS

TĀLIVALDIS IVARS ŠMITS
PRIMARY EXAMINER